Computing for the DOE Genomes to Life Program

Ed Uberbacher[3]
David Thomassen[1]
Ari Patrinos[1]
Gary Johnson[2]
Carl E. Oliver[2]
Marvin Frazier[1]
[1]*Office of Science, Office of Biological and Environmental Research*
[2]*Office of Science, Office of Advanced Scientific Computing Research*
[3]*Oak Ridge National Laboratory*

Abstract

A key goal of the DOE Office of Science's systems biology program, Genomes to Life (GTL), is to achieve, over the next 10 to 20 years, a basic understanding of thousands of environmental microbes and microbial systems in their native environments. This goal demands that we develop new models for scientific discovery that integrate methods and systems which tightly couple advanced computing, mathematics, algorithms, and data-management technologies with large-scale experimental data generation.

Key GTL computing challenges include (1) providing data management and user access for the immensely complex data from microbial biological systems, ranging from comprehensive proteomics measurements, to molecular machines and interaction networks, regulatory networks and metabolic pathways, (2) developing new types of analysis methods and strategies for unprecedented levels of comprehensive data reduction from many different high-throughput technologies such as mass spectrometry and imaging, (3) constructing models and simulating many aspects of biological systems, including molecular interactions and dynamics, metabolic, regulatory, and signaling pathways and networks, and dynamic cell models, and (4) implementing the computing infrastructure and middleware to accomplish large-scale distributed analysis, simulations on massively parallel architectures, and achieving effective access to petabytes of biological data.

The vast amount of information, contained in the hundreds of genomes sequenced and the thousands to come, offers an unprecedented opportunity to capture the behavior of complex microbial systems in mathematical and computational form. This will provide a new foundation for solving urgent problems in energy security, global climate change mitigation and environmental remediation.